

## Durham Research Online

---

### Deposited in DRO:

19 September 2016

### Version of attached file:

Published Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Wang, Xing (2013) 'The rate function of S-mixing process and its application in evaluating empirical likelihood tests.', *Journal of mathematics and statistics.*, 9 (4). pp. 334-338.

### Further information on publisher's website:

<http://dx.doi.org/10.3844/jmssp.2013.334.338>

### Publisher's copyright statement:

© 2013 Xing Wang. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Additional information:

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# THE RATE FUNCTION OF S-MIXING PROCESS AND ITS APPLICATION IN EVALUATING EMPIRICAL LIKELIHOOD TESTS

Xing Wang

Department of Economics and Finance, Business School, 23-26 Old Elvet, Durham University, DH1 3HY, UK

Received 2013-03-28; Revised 2013-07-09; Accepted 2013-10-28

## ABSTRACT

In this study we show that the rate function of S-Mixing stationary process with part of the hypermixing condition is equivalent to the Kullback-Leibler distance. This finding can be used to extend the result of the optimality of empirical likelihood test in i.i.d setting to weakly depended case, so that the Asymptotic Relative Efficiency (ARE) of the empirical likelihood test with strong mixing data can be established.

**Keywords:** Large Deviations, Strong Mixing, Empirical Likelihood Test

## 1. INTRODUCTION

Large deviation results on random variables can be used to derive the bounds of some loss functions, therefore the efficiency of statistical inference can be assessed by large deviation probabilities, particularly in the case that both the loss function and the rate function of the large deviations are the Kullback-Leibler (K-L) distance between two probability measure, say,  $m_1$  and  $m_2$ :

$$H(m_1 \| m_2) = \begin{cases} \int \log \left( \frac{dm_1}{dm_2} \right) dm_1, & \text{if } m_1 \ll m_2 \\ \infty, & \text{otherwise} \end{cases}$$

Bahadur (1967) and Dembo and Zeitouni (1998) (hereafter DZ), where the optimality of statistical inference with i.i.d data is extensively discussed. Although in general the rate function for non i.i.d. data is not equal to the K-L distance, the purpose of this note is to show that combined with part of the hypermixing condition, the S-mixing stationary process has a rate function equal to the K-L distance, hence the optimality results on empirical likelihood test with i.i.d. data (Kitamura, 2001) can be extended to the case of weakly dependent data.

Let  $\Sigma$  be a compact topological space and  $\{X_t; t \in \mathbb{Z}\}$  be a stationary stochastic process taking values in  $\Sigma$ .

Also let  $\mathcal{F}_a^b = \sigma(X_i : a \leq i \leq b)$  denote the  $\sigma$ -algebra generated by  $\{X_i; a \leq i \leq b\}$ . According to Bryc and Dembo (1996) (hereafter BD),  $\{X_t; t \in \mathbb{Z}\}$  is said to be S-mixing if for any finite constant  $C < \infty$ , there exists a non-decreasing sequence  $l(n) \in \mathbb{N}$  with Equation 1:

$$\sum_{n=1}^{\infty} \frac{l(n)}{n(n+1)} < \infty \quad (1)$$

Such that Equation 2:

$$\sup |P(A)P(B) - e^{l(n)}P(A \cap B)| \leq e^{-Cn} \quad (2)$$

where,  $A \in \mathcal{F}_0^{k_1}, B \in \mathcal{F}_{k_1+l(n)}^{k_1+k_2+l(n)}$ ,  $k_1, k_2 \in \mathbb{Z}_+$ .

BD points out that S-mixing is a fairly weak condition since  $\alpha$ -mixing condition suffices for S-mixing, so it is suitable for a quite general class of stochastic processes (see also DZ). Specifically, BD shows that  $\alpha$ -mixing implies S-mixing and also proves that S-mixing will hold if it satisfies the following two conditions (H-1) and (H-2), which are sometimes called hypermixing conditions.

### Assumption 1 (H-1)

There exist  $l, \alpha < \infty$  such that, for all  $k, r < \infty$  and any  $l$ -separated functions (For any given integers  $r \geq k \geq 2, l$

$\geq 0$ , a family of functions  $\{f_i\}_{i=1}^k \in B(\Sigma^r, \mathbb{R})$  is called  $l$ -separated if there exist  $k$  disjoint intervals  $I_1, \dots, I_r$ , such that  $\text{dist}(I_m, I_{m'}) \geq l$  for  $1 \leq m < m' \leq r$  and  $f_m$  is  $I_m$  measurable for any  $1 \leq m < n$   $f_i \in B(\Sigma, \mathbb{R})$  Equation 3:

$$E \left| \prod_{i=1}^k f_i(X_1, \dots, X_r) \right| \leq \prod_{i=1}^k \left( E \left[ |f_i(X_1, \dots, X_r)|^\alpha \right] \right)^{1/\alpha} \quad (3)$$

### Assumption 2 (H-2)

There is some constant  $l_0$  and functions.  $\beta(l) \geq 1$ ,  $\omega(l) \geq 0$  such that for all  $l > l_0$ , all  $r < \infty$  and any two  $l$ -separated functions  $f, g \in B(\Sigma^r)$  Equation 4:

$$\begin{aligned} & \left| E[f(X_1, \dots, X_r)] E[g(X_1, \dots, X_r)] - E[f(X_1, \dots, X_r)g(X_1, \dots, X_r)] \right| \\ & \leq \omega(l) \left( E \left[ |f(X_1, \dots, X_r)|^{\beta(l)} \right] \right)^{1/\beta(l)} \\ & \quad \left( E \left[ |g(X_1, \dots, X_r)|^{\beta(l)} \right] \right)^{1/\beta(l)} \end{aligned} \quad (4)$$

The following theorem of large deviations can be considered as an analogue of Sanov theorem (Sanov, 1965) of  $S$ -mixing process, with which we are able to evaluate large deviation probabilities of weakly dependent data. Let  $Q$  be the underlying probability measure of the whole process and let  $Q_n$  denote the  $n$ th margin of  $Q$  and particularly,  $Q_1 \in M_1(\Sigma)$  is the probability measure of a single realization, where  $M_1(\Sigma)$  denotes the space of Borel probability measures on  $\Sigma$ . Also define the empirical measure of a sample

$$x_1, x_2, \dots, x_n \text{ to be } \mu_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{x_i \leq x}.$$

### Theorem 1 (Bryc and Dembo, 1996)

If a stationary process  $\{X_t: t \in \mathbb{Z}\}$  is  $S$ -mixing, the sequence of empirical measures satisfies the LDP with respect to the  $\tau$ -topology in  $M_1(\Sigma)$  and this LDP is governed by the good rate function Equation 5:

$$I(u) = \sup_{f \in B(\Sigma, \mathbb{R})} \left\{ \int_{\Sigma} f d\mu - \Lambda(f) \right\} \quad (5)$$

i.e., for every set  $\Gamma \subseteq M_1(\Sigma)$  Equation 6 and 7:

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P^n(\mu_n \in \Gamma) \geq - \inf_{\mu \in \Gamma^0} I(\mu) \quad (6)$$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n(\mu_n \in \Gamma) \leq - \inf_{\mu \in \Gamma} I(\mu) \quad (7)$$

where,  $P^n \in M_1(\Sigma^n)$  is the distribution of  $\mu_n$  and Equation 8:

$$\Lambda(f) = \lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_n(f) = \lim_{n \rightarrow \infty} \frac{1}{n} \log E_Q \left[ \exp \left( \sum_{i=1}^n f(x_i) \right) \right] \quad (8)$$

and the limit exists for every  $f \in B(\Sigma, \mathbb{R})$ , where  $B(\Sigma, \mathbb{R})$  is the space of all bounded, real-valued, Borel measurable functions on  $\Sigma$ .

On the rate function of the LDP, BD mentioned roughly in their paper that  $I(v)$  in general will be less than specific K-L distance, but they didn't provide any proof. However, we find that  $I(v)$  will be equal to the K-L distance if the  $S$ -mixing condition are combined with part of hypermixing condition. This main result is presented in the next section.

## 2. THE RATE FUNCTION OF THE S-MIXING PROCESS

First, we introduce the following lemma from DZ.

### Lemma 1

Given assumption (H-1), for  $\gamma > 0$ , we have Equation 9:

$$\Lambda(f) \leq \frac{1}{\gamma} \log \int_{\Sigma} e^{\gamma f(x)} dQ_1 \quad (9)$$

see DZ for proof. This lemma ensures that  $\Lambda(f)$  is bounded and so is its Legendre transformation  $I(v)$  in (5).

### Assumption 3

If  $v \ll Q_1$ , then the Radon-Nikodym density  $dv/dQ_1$  is bounded.

Our following theorem shows the equivalence of the Kullback-Leibler distance  $H(\cdot \| Q_1)$  and the rate function of the LDP of  $S$ -mixing data.

### Theorem 2

If assumption (H-1) and 3 are satisfied, the rate function  $I(v)$  in theorem 1 satisfies:

$$I(v) = H(v \| Q_1)$$

### Proof

Firstly we show that  $I(v) \geq H(v \| Q_1)$ : From lemma 1 we have Equation 10:

$$\begin{aligned}
I(v) &= \sup_{f \in B(X, \mathbb{R})} \left\{ \int_{\Sigma} f d\mu - \Lambda(f) \right\} \\
&\geq \sup_{f \in B(X, \mathbb{R})} \left\{ \int_{\Sigma} f d\mu - \frac{1}{\gamma} \log \int_{\Sigma} e^{\gamma f(x)} dQ_1 \right\} \\
&\geq \int_{\Sigma} f d\mu - \frac{1}{\gamma} \log \int_{\Sigma} e^{\gamma f(x)} dQ_1
\end{aligned} \quad (10)$$

The last inequality implies that  $v$  is absolute continuous with respect to  $Q_1$ . To see this, let  $\Gamma \in A$  satisfying  $Q_1(\Gamma) = 0$ . Because the inequality holds for any  $f \in B(X, \mathbb{R})$ , we can take  $f = \xi I_{\Gamma}$  where  $\xi > 0$  and  $I$  is the indicator function. Note that  $\int_{\Sigma} e^{\gamma \xi I_{\Gamma}} dQ_1 = 1$ , so we have  $I(v) \geq \xi v(\Gamma)$  for any  $\xi > 0$ . Since  $I(v)$  is non-negative, we conclude  $v(\Gamma) = 0$ , i.e.,  $v \ll Q_1$ . Therefore the Radon-Nikodym derivative of  $v$  with respect to  $Q_1$  exists, namely,  $\phi \equiv \frac{dv}{dQ_1}$ . Hence  $I(v) \geq H(v \| Q_1)$  is implied by (10) if we take  $f = \log \phi$  with assumption 3.

On the other hand, by Jensen's inequality and the stationarity of  $\{x_i\}$ , we obtain:

$$\begin{aligned}
\Lambda(f) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log E \left[ \exp \left( \sum_{i=1}^n f(x_i) \right) \right] \\
&\geq \lim_{n \rightarrow \infty} \frac{1}{n} E \left[ \log \exp \left( \sum_{i=1}^n f(x_i) \right) \right] \\
&= E[f(x)] \\
&\geq \int_{\Sigma} f(x) dx - H(v \| Q_1)
\end{aligned}$$

which completes the proof.

In the next section we use Theorem 2 to establish the asymptotic optimality of the empirical likelihood (EL) test in a general Neyman-Pearson sense, showing the asymptotic type II error probability can achieve the lower bounds indicated by the rate function of the large deviation results. Similar results with i.i.d data are shown by Zeitouni and Gutman (1991) (hereafter ZG) and Kitamura (2001), but our analysis will be in a context of weakly dependent data.

### 3. THE OPTIMALITY OF THE EL TEST

#### 3.1. The EL Test Statistic

Let  $\{x_i\}_{i=1}^n$  be a realization of a stationary  $\alpha$ -mixing (and hence ergodic and S-mixing) process  $\{X_t; t \in \mathbb{Z}\}$  taking values on  $\Sigma$ . We are interested in applying the EL to test the following moment condition Equation 11:

$$E[g(x_i, \theta_0)] = \int_{\Sigma} g(x_i, \theta_0) dQ_i = 0 \quad (11)$$

where the moment indicator  $g: \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^m$  is continuous for all  $d$ -dimensional  $x_i \in \Sigma$  and  $Q_i$  is the unknown distribution of  $x_i$ , i.e.,  $Q_i$  is the marginal of  $Q$  at  $x_i$ . Also  $\theta_0 \in \Theta \subset \mathbb{R}^p$  is the true parameter vector. We consider the over-identifying case where  $m \geq p$ .

If we assign each observation  $x_i$  with a probability  $p_i$ , the EL method solves the following problem:

$$\begin{aligned}
&\max \prod_{i=1}^n p_i \\
&\text{s.t. } \sum_{i=1}^n p_i g(x_i, \theta_0) = 0, \sum_{i=1}^n p_i = 1, p_i \geq 0
\end{aligned}$$

Qin and Lawless (1994). The test statistic is given by Kitamura (1997) using block technique Equation 12:

$$W = \inf_{\theta \in \Theta} \sup_{\lambda \in \mathbb{R}^m} 2 \left( \frac{n}{TM} \right) \sum_{i=1}^T \log(1 + \lambda' b(B_i, \theta)) \xrightarrow{d} X_{m-p}^2 \quad (12)$$

where  $\lambda$  is a vector of Lagrangian multipliers and Equation 13:

$$b(B_i, \theta) = \frac{1}{M} \sum_{j=1}^M g(x_{(i-1)L+j}, \theta) \quad (13)$$

where,  $T$  is the number of blocks,  $M > 1$  denotes the block length,  $L$  is the separation between block starting points.

Now define Equation 14:

$$Q(\theta) = \left\{ \mu \in M_1(\Sigma) : \int g(x, \theta) d\mu = 0 \right\} \quad (14)$$

Let  $Q = \bigcup_{\theta \in \Theta} Q(\theta)$  thus  $Q$  is the set of probability measures which satisfy the moment condition over the parameter space. Hence the problem of testing (11) can be considered as testing Equation 15:

$$H_0: \mu_n \in Q \text{ versus } H_1: \mu_n \notin Q \quad (15)$$

where,  $\mu_n$  is the empirical measure of the observed data. Intuitively, empirical likelihood test is to investigate whether the empirical measure  $\mu_n$  which is constructed to be as close to the true probability measure as possible by the EL method, is too far away from any of the measures in  $Q$  or not. Therefore, considering the K-L distance

$H(\cdot, \cdot)$  as a measure of distance between two probability measures, the EL test statistic  $W$  is indeed a result of the following minimizing problem Equation 16:

$$\inf_{Q_1 \in Q} H(\mu_n \| Q_1) \quad (16)$$

consequently, the empirical likelihood ratio test is to reject  $H_0$  if Equation 17:

$$\inf_{Q_1 \in Q} H(\mu_n \| Q_1) > c \quad (17)$$

for some threshold constant  $c > 0$ . This is to say, under the null hypothesis, the empirical measure  $\mu_n \in Q$  and therefore, if the distance between  $\mu_n$  and any of the probability measures in  $Q$  is too large, then we shall reject the null hypothesis. It also tells that the test depends on the data only through  $\mu_n$  (see DZ). Thus empirical likelihood test can be considered as a sequence of partitions  $\Lambda(n) = (\Lambda_1(n), \Lambda_2(n))$  of  $M_1(\Sigma)$  where  $n = 1, 2, \dots$  and Equation 18:

$$\Lambda_1 = \left\{ \mu \in M_1(\Sigma) : \inf_{Q_1 \in Q} H(\mu \| Q_1) < c \right\}, \Lambda_2 = M_1(\Sigma) \setminus \Lambda_1 \quad (18)$$

In the following we abbreviate  $(\Lambda_1(n), \Lambda_2(n))$  as  $(\Lambda_1, \Lambda_2)$  for economy of notation, but its dependence on the sample size  $n$  should not be ignored. Since in general framework, pointwise bounds on error probabilities are not available (Kitamura, 2001), we consider the  $\delta$ -smoothing of the set  $\Lambda_2$ :

$$\Lambda_2^\delta = \bigcup_{\mu \in \Lambda_2} B(\mu, \delta)$$

and:

$$\Lambda_1^\delta = M_1(\Sigma) \setminus \Lambda_2^\delta$$

where,  $B(\mu, \delta)$  denotes an open ball of radius  $\delta$  around  $\mu$  and the balls are taken in the Levy metric:

$$d(\mu_1, \mu_2) = \inf\{\epsilon > 0 : \mu_1(A) \leq \mu_2(A) + \epsilon \forall A \in \mathcal{A}\}$$

which is compatible with the weak, strong and uniform convergence of discrete probability measures (e.g., see ZG).

### 3.2. Optimality Argument

To directly apply large deviation property of  $\mu_n$  in Theorem 1 and Theorem 2 to establish the optimality of

the EL test, firstly we need some tightness and continuous condition.

#### Assumption 4

$\sup_{\theta \in \Theta} \|g(x, \theta)\|$  is bounded almost surely and thus it is a random variable under all  $Q_1 \in Q$ . The functional  $\inf_{Q_1 \in Q} H(\mu_n \| Q_1)$  is uniformly continuous in  $\mu \in M_1(\Sigma)$  in the  $\tau$ -topology.

#### Lemma 2

$$\bar{\Lambda}_1 = \left\{ \mu \in M_1(\Sigma) : \inf_{Q_1 \in Q} H(\mu \| Q_1) < c \right\}$$

#### Proof

Since  $H(\mu \| Q_1)$  is a lower-semicontinuous function, the set  $\left\{ \mu \in M_1(\Sigma) : \inf_{Q_1 \in Q} H(\mu \| Q_1) \leq c \right\}$  is closed. So we have Equation 19:

$$\bar{\Lambda}_1 \subseteq \left\{ \mu \in M_1(\Sigma) : \inf_{Q_1 \in Q} H(\mu \| Q_1) \leq c \right\} \quad (19)$$

To see the other direction, notice that assumption 4-b implies that  $\mu \in \{\mu : \inf_{Q_1 \in Q} H(\mu \| Q_1) = c\}$  is a limit point of  $\Lambda$ . Hence the lemma follows.

In the next theorem we show that EL test is the uniformly most powerful test among all the tests with the same size. This uniform optimality is sometimes called universal property in information theory, e.g., see ZG and DZ.

#### Theorem 3

Let  $P_i^n = \otimes_{n=1}^n \mu \in M_1(\Sigma^n)$  with  $i = 0, 1$ , be the law of the empirical measure under the hypothesis  $H_0$  and  $H_1$  respectively. Then the empirical likelihood test  $(\Lambda_1, \Lambda_2)$  satisfies Equation 20:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_1^n \left\{ \mu_n \in \Lambda_2^\delta \right\} \leq -c \quad (20)$$

Moreover, for any another test  $(\Omega_1, \Omega_2)$  which is also a partition of  $M_1(\Sigma)$  and satisfies:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_1^n \left\{ \mu_n \in \Omega_2^\delta \right\} \leq -c$$

we have Equation 21:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_2^n \{ \mu_n \in \Omega_1 \} \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log P_2^n \{ \mu_n \in \Lambda_1 \} \quad (21)$$

### Proof

The inequality (19) implies that the asymptotic type I error probability of the EL test is bounded above from  $e^{-nc}$ . This boundness is straightforward from the LDP of the empirical measure  $\mu_n$  indicated in theorem 1 and theorem 2 given assumption 4:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_1^n \{ \mu_n \in \Lambda_2 \} \leq - \inf_{u \in \Lambda_2} I(u) = - \inf_{u \in \Lambda_2} H(v \| Q_1) \leq -c$$

Now we show that there exists some  $n_0 \in \mathbb{N}$ , such that  $\Lambda_1 \subseteq \Omega(n)$  for all  $n > n_0$  along the limit supremum. Suppose it is not so. Then there exists a subsequence  $n_k$  such that  $\omega_{n_k} \in \bar{\Lambda}_1$  and  $\omega_{n_k} \in \Omega_2^\delta$ . Since the set  $\bar{\Lambda}_1$  is compact due to lemma 2, there exists some  $\omega \in \bar{\Lambda}_1$  such that  $\omega_{n_k} \rightarrow \omega$ . Note that  $\omega_{n_k} \in \Omega_2^\delta$ , thus  $B(\omega_{n_k}, \delta) \subset \Omega_2^\delta(n_k)$  and  $B(\omega, \delta/2) \subset \Omega_2^\delta(n_k)$  hold for infinitely many  $n_k$ . So:

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{n} \log P_2^n \{ \mu_n \in \Omega_1 \} \\ & \geq \liminf_{n \rightarrow \infty} \frac{1}{n_k} \log P_2^{n_k} \{ \mu_{n_k} \in \Omega_2^\delta(n_k) \} \\ & \geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_2^n \{ \mu_n \in B(\omega, \delta/2) \} \\ & \geq \inf_{u \in B(\omega, \delta/2)} I(u) \\ & = - \inf_{u \in B(\omega, \delta/2)} H(v \| Q_1) \\ & \geq -c \end{aligned}$$

this contradicts (20) and thus  $\Lambda_1^\delta \subseteq \Omega_1^\delta$  is verified. Consequently:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_2^n \{ \mu_n \in \Omega_1 \} \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log P_2^n \{ \mu_n \in \Lambda_1 \}.$$

## 4. CONCLUSION

We have shown the equivalence of the rate function of the S-mixing process to the K-L distance if part of the hypermixing condition is added. This results can be applied to assess the efficiency of statistical inference based on minimum distance method in weakly dependent setting, where two probability measureds can be compared by the K-L distance.

## 5. REFERENCES

- Bahadur, R.R., 1967. Rates of Convergence of Estimates and Test Statistics. *Ann. Math. Statist.*, 38: 303-324. DOI: 10.1214/aoms/1177698949
- Bryc, W. and A. Dembo, 1996. Large deviations and strong mixing. *Ann. l'I.H.P. Probab. Statist.*, 32: 549-569.
- Dembo, A. and O. Zeitouni, 1998. *Large Deviations Techniques and Applications*. 2nd Edn., Springer, New York, Berlin, ISBN-10: 3642033113, pp: 412.
- Kitamura, Y., 1997. Empirical likelihood methods with weakly dependent processes. *Ann. Statist.*, 25: 2084-2102.
- Kitamura, Y., 2001. Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, 69: 1661-1672. DOI: 10.1111/1468-0262.00261
- Qin, J. and J. Lawless, 1994. Empirical likelihood and general estimating equations. *Ann. Statist.*, 22: 300-325. DOI: 10.1214/aos/1176325370
- Sanov, I.N., 1965. On the probability of large deviations of random variables. *Selected Trans. Mathem. Statist. Probab. I*.
- Zeitouni, O. and M. Gutman, 1991. On universal hypotheses testing via large deviations. *IEEE Trans. Inform. Theory*, 37: 285-290. DOI: 10.1109/18.75244